

CS130 Assignment #4

Date assigned: Monday, November 2, 2009

Date due: Monday, November 9, 2009

Points: 50

Create a Word document **PUNuetIDAnswers.doc** using your PUNetID that will contain answers to each of the following questions in order when applicable. Your answer document and any other files created are to be placed in a folder called **PUNetIDSolution4** once again using your PUNetID.

Problem #1

Consider the calories burned by women and men over a 24-hour period of time. Using SPSS, you are to create a dataset called **Calories.sav** with the data below.

Subject	Sex	Rate
1	M	1788
2	M	1676
3	F	994
4	F	1431
5	M	1400
6	F	1413
7	M	1358
8	F	1503
9	M	1260
10	M	1615
11	F	1190
12	F	915
13	M	1461
14	F	1125
15	M	1060
16	F	1350
17	F	1210
18	M	1870

In the Word document described above, answer each of the following questions in order under the heading specified. That is, list the heading in bold and then give your answer.

You are to find if the calories burned by women is significantly different than the calories burned by men.

Problem1 – Question 1

What hypothesis test will you need to perform to find this result? Fully explain why you selected the test you did.

Problem1 – Question 2

State the NULL Hypothesis for the statistical test you selected.

Problem1 – Question 3

Paste in the results of the statistical test.

Problem1 – Question 4

State your conclusion.

Problem1 – Question 5

Explain exactly what the Sig. (2-tailed) value means in this case. I don't want just if the value is less than 0.05 reject the null hypothesis or greater than 0.05 accept the null hypothesis, but in terms of the problem explain the meaning of the Sig. (2-tailed) value including the concept of chance occurrences. Someone without a statistics background must be able to understand your explanation.

Problem1 – Question 6

What are the correct type and measure of each of the variables in this dataset? Explain why each variable is the type and measure you selected.

Problem #2

We sometimes use a correlation analysis to indicate the degree of a linear relationship that exists between two variables. Typically the researcher will gather two or more observed values from each subject (or sampling unit) and then try and see if there is a linear relationship between any of the variables. SPSS computes the Pearson correlation coefficient which has an absolute value between 0 and 1 where a 1 indicates a perfect linear relationship and a 0 means that no linear relationship exists whatsoever. If two variables increase or decrease proportionately, then we say a positive correlation exists between these two variables. If one variable increases while another decreases, then there is a negative correlation.

In class, we will grab the Lipid dataset and produce the correlation matrix of Cholesterol, Triglycerides, HDL, and LDL. We will identify and discuss the strongest correlation given in the matrix produced by the variables mentioned.

In the CS130 Public folder on Turing is a data file called WesternStates.txt that is a "tab" delimited data file. The file consists of scores for several different variables where the higher the score, the better, except for the variables Housing and Crime. In these two cases, the lower the score, the better.

Create an SPSS dataset from the WesternStates.txt by importing the data into SPSS as we have done in class several times. Save the dataset as **WesternStates.sav**.

In the same Word document from Problem 1, report on each of the following:

Problem2 – Question 1

What two different variables have the highest correlation and what is the Pearson correlation coefficient?

Problem2 – Question 2

Create and paste into your Word document a Scatterplot of these two variables produced by SPSS.

Problem2 – Question 3

Perform a linear regression on the two different variables with the highest correlation and paste in the Coefficients table produced by SPSS.

Problem2 – Question 4

Using the Equation Editor and the Coefficient table, write out the linear regression equation and R^2 value.

Problem2 – Question 5

Perform the same linear regression using Excel and paste in a graph of the linear regression where the graph includes the linear equation and R^2 value.

Note1: Make sure all of your graphs and charts are properly and accurately labeled. Also, make sure they look professional in the Word document.

Note2: The Word document answers are to be in the exact order as specified above. As an example, do not put Step #6 results ahead of Step #2 results in the Word document.

Submitting your work:

To submit your work, copy your single folder **PUNetIDAssign4** (for me that's ryandjAssign4) containing (**PUNetIDAnswers.doc**, **Calories.sav**, and **WesternStates.sav**) into the **CS130 Drop** folder on Turing. You must submit your work by the time specified above for your assignment to be considered on time. Reread the syllabus for the late policy. The official time can be found <http://time.gov/timezone.cgi?Pacific/d/-8/java>.

Grading:

Grading will be based on:

- 1) Correctness of your results
- 2) Completeness of your results
- 3) Professional look and correctness of your Word document answers.

Be sure to come see me early with any questions! Also, make sure you reread the Academic Dishonesty policy from the course syllabus. This is NOT a group project.